

PENERAPAN *A PRIORI* RULE UNTUK MELIHAT PERILAKU PENGGUNAAN INTERNET

Artikel Ilmiah



Peneliti:

**Robertho F Wellikin (672011203)
Dr. Irwan Sembiring, S.T., M.Kom.
Ramos Somya, S.Kom., M.Cs.**

**Program Studi Teknik Informatika
Fakultas Teknologi Informasi
Universitas Kristen Satya Wacana
Salatiga
Mei 2016**



PERPUSTAKAAN UNIVERSITAS
UNIVERSITAS KRISTEN SATYA WACANA
Jl. Diponegoro 52 – 60 Salatiga 50711
Jawa Tengah, Indonesia
Telp. 0298 – 321212, Fax. 0298 321433
Email: library@adm.uksw.edu ; http://library.uksw.edu

PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Robertho Franklin Wellikin
NIM : 672011203 Email : 672011203@student.uksw.edu
Fakultas : Fakultas Teknologi Informasi Program Studi : Teknik Informatika
Judul tugas akhir : Penerapan *A Priori Rule* Untuk Melihat Perilaku Pengguna Internet

Pembimbing : 1. Dr. Irwan Sembiring, S.T., M.Kom.
2. Ramos Somya, S.Kom., M.Cs.

Dengan ini menyatakan bahwa:

1. Hasil karya yang saya serahkan ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar kesarjanaan baik di Universitas Kristen Satya Wacana maupun di institusi pendidikan lainnya.
2. Hasil karya saya ini bukan saduran/terjemahan melainkan merupakan gagasan, rumusan, dan hasil pelaksanaan penelitian/implementasi saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing akademik dan narasumber penelitian.
3. Hasil karya saya ini merupakan hasil revisi terakhir setelah diujikan yang telah diketahui dan disetujui oleh pembimbing.
4. Dalam karya saya ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali yang digunakan sebagai acuan dalam naskah dengan menyebutkan nama pengarang dan dicantumkan dalam daftar pustaka.

Pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari terbukti ada penyimpangan dan ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya saya ini, serta sanksi lain yang sesuai dengan ketentuan yang berlaku di Universitas Kristen Satya Wacana.

Salatiga, 27 Juni 2016

Robertho Franklin Wellikin

F-LIB-080



PERNYATAAN PERSETUJUAN AKSES

Saya yang bertanda tangan di bawah ini:

Nama : Robertho Franklin Wellikin
NIM : 672011203 Email : 672011203@student.uksw.edu
Fakultas : Fakultas Teknologi Informasi Program Studi : Teknik Informatika
Judul tugas akhir : Penerapan *A Priori Rule* Untuk Melihat Perilaku Pengguna Internet

Dengan ini saya menyerahkan hak *non-eksklusif** kepada Perpustakaan Universitas – Universitas Kristen Satya Wacana untuk menyimpan, mengatur akses serta melakukan pengelolaan terhadap karya saya ini dengan mengacu pada ketentuan akses tugas akhir elektronik sebagai berikut (beri tanda pada kotak yang sesuai):

- ☒ a. Saya mengizinkan karya tersebut diunggah ke dalam aplikasi Repositori Perpustakaan Universitas, dan/atau portal GARUDA
- ☐ b. Saya tidak mengizinkan karya tersebut diunggah ke dalam aplikasi Repositori Perpustakaan Universitas, dan/atau portal GARUDA**

* Hak yang tidak terbatas hanya bagi satu pihak saja. Pengajar, peneliti, dan mahasiswa yang menyerahkan hak non-eksklusif kepada Repositori Perpustakaan Universitas saat mengumpulkan hasil karya mereka masih memiliki hak copyright atas karya tersebut.

** Hanya akan menampilkan halaman judul dan abstrak. Pilihan ini harus dilampiri dengan penjelasan/ alasan tertulis dari pembimbing TA dan diketahui oleh pimpinan fakultas (dekan/kaprodi).

Demikian pernyataan ini saya buat dengan sebenarnya.

Salatiga, 27 JUNI 2016

1956

Mengetahui,

Robertho Franklin Wellikin

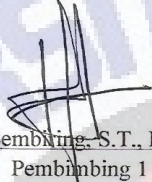
Dr. Irwan Sembiring, S.T., M.Kom.


Ramos Somya, S.Kom., M.Cs.

Lembar Pengesahan

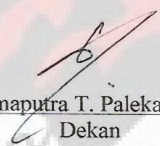
Judul Tugas Akhir : Penerapan *A Priori Rule* Untuk Melihat Perilaku Pengguna Internet
Nama Mahasiswa : Robertho F Wellikin
NIM : 672011203
Program Studi : Teknik Informatika
Fakultas : Teknologi Informasi

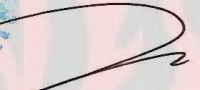
Menyetujui,


Dr. Irwan Sembiring, S.T., M.Kom.
Pembimbing 1


Ramos Somya, S.Kom., M.Cs.
Pembimbing 2

Mengesahkan,

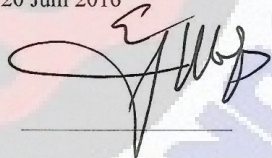
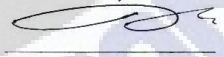

Dr. Dharmaputra T. Palekahelu, M.Pd.
Dekan


Supriyadi, S.Si., M.Kom.
Ketua Program Studi

Dinyatakan Lulus Ujian tanggal: 20 Juni 2016

Penguji:

1. Prof. Dr. Ir. Eko Sedyono, M.Kom.
2. Hindriyanto D. Purnomo, Ph.D.

1. Pendahuluan

Penggunaan internet saat ini bukan lagi sesuatu yang baru di berbagai lapisan masyarakat dengan berbagai kebutuhan akan informasi. Universitas Kristen Satya Wacana (UKSW) Salatiga menyediakan sarana akses internet yang terpusat pada pusat data yang dikonsumsi secara langsung oleh setiap mahasiswa untuk berbagai keperluan akademik maupun non akademik.

Algoritma apriori adalah algoritma pengambilan data dengan aturan asosiatif (Association rule) untuk menentukan hubungan asosiatif suatu kombinasi item [2]. Association Rule yang dimaksud dilakukan melalui mekanisme penghitungan *support* dan *confidence* dari suatu hubungan item. Sebuah rule asosiasi dikatakan interesting jika nilai support adalah lebih besar dari minimum support dan juga nilai confidence adalah lebih besar dari minimum confidence. Algoritma apriori ini akan cocok untuk diterapkan bila terdapat beberapa hubungan item yang ingin dianalisa. Salah satunya yang bisa diterapkan adalah pada perilaku pengguna internet.

Permasalahan yang terjadi saat ini adalah belum adanya metode yang mampu menghasilkan rule untuk memprediksi situs - situs yang paling banyak dikunjungi oleh mahasiswa, maka diperlukan rule yang mampu melihat perilaku penggunaan internet setiap harinya pada waktu – waktu tertentu apa saja yang diakses pengguna akses.

Salah satu metode yang diharapkan mampu mengatasi permasalahan tersebut adalah A Priori atau dikenal dengan *association rule*. A Priori adalah salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) untuk menghasilkan algoritma yang efisien.

Dalam penelitian ini digunakan data log yang diambil dari server Universitas Kristen Satya Wacana (UKSW) salatiga, dimana pada data log kita bisa lihat apa saja yang diakses oleh para pengguna internet. Setelah itu kita tentukan *variable* yang cocok pada data log agar kita bisa mudah terapkan apriori rule. Penerapan A Priori rule digunakan untuk mengetahui perilaku penggunaan internet, rule yang dihasilkan mampu memprediksi situs mana yang paling sering diakses pada periode tertentu. Dengan mengetahui rule atau pola yang dihasilkan, maka pihak pengelola internet dapat mengambil keputusan dalam penyediaan bandwidth pada waktu – waktu tertentu.

2. Kajian Pustaka

Penelitian-penelitian yang telah dilakukan sebelumnya dijadikan sebagai dasar atau pembandingan dalam merancang algoritma. Penelitian pertama dengan topik “ *Implementasi Data Mining Algoritma A Priori Pada Sistem Persediaan Alat-alat Kesehatan* ” yang dilakukan oleh Kennedy Tampubolon[1]. Dalam penelitian ini, peneliti telah membaca banyak hal tentang data mining, teknik data mining dengan association rules, dan algoritma – algoritma yang terkait dengan *association rules* serta implementasinya dalam bisnis. Peneliti ingin mengimplementasikan teknik data mining pada database barang terjual alat – alat kesehatan dengan aturan asosiasi menggunakan algoritma *Apriori* untuk mengetahui pola kombinasi item dan itemset frekuensi tinggi dari alat – alat kesehatan sehingga dapat dijadikan sebagai faktor pengambilan keputusan dalam memprediksi persediaan barang. Untuk itu penelitian memilih

Apotek Kelambir – 2 Medan sebagai tempat riset dan pengambilan data sampel untuk diolah dan dijadikan objek dalam pengujian.

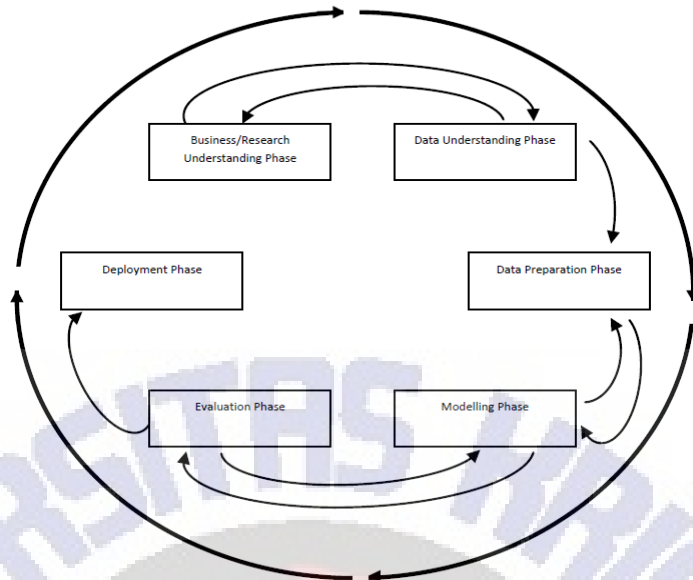
Penelitian dilakukan oleh Robi Yanto[2] dengan topik “*Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat*”. Penelitian ini melakukan analisa data dengan menggunakan data mining dan metode algoritma apriori. Sistem yang dibangun ditujukan untuk pemenuhan dalam penentuan pola pembelian obat dengan menggunakan bahasa pemrograman Visual Basic 6.0 dan database Mysql pada studi kasus di sektor kesehatan. Sistem ini dibangun berdasarkan kebutuhan pengguna yang diperoleh melalui metode wawancara dan studi lapangan. Metodologi pengembangan sistem yang digunakan yaitu metode waterfall yang terdiri Analisis, Desain, Pengkodean dan Pengujian. Hasil pengujian dengan algoritma apriori dan sistem yang dibangun menunjukkan hasil yang telah memenuhi kebutuhan dalam penentuan pola pembelian obat berdasarkan kecenderungan pembelian obat oleh pelanggan. Dibandingkan dengan sistem yang sedang berjalan kinerja tersebut ditunjukkan pada efektifitas informasi dari sistem tentang penentuan pola pembelian obat untuk ketersediaan obat dan tata letak obat untuk memudahkan dalam mengetahui keberadaan obat yang dilihat dari 2 itemset obat. Penelitian ini dijadikan sebagai acuan untuk menentukan pola dari data yang sudah kita dapat sehingga kita bisa melihat perilaku pengguna internet. Untuk mendapat pola kita perlu menentukan variable atau item yang diakses bersamaan sehingga kita dapat menerapkan algoritma *A Priori Rule*.

Data mining merupakan disebut sebagai proses ekstraksi pengetahuan dari data yang besar. Sesuai fungsinya, data mining adalah proses pengambilan pengetahuan dari volume data yang besar yang disimpan dalam basis data, data *warehouse*, atau informasi yang disimpan dalam repository [1]. Gartner Group dalam [1], menyebutkan bahwa data mining adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repository atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan teknik matematika.

Menurut Larose [2], data mining merupakan proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repository, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika.

Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki [2].

Data mining adalah sebuah proses, sehingga dalam melakukan prosesnya harus sesuai prosedur yaitu proses CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yaitu sebagai keseluruhan proses, preproceasing data, pembentukan model, model evaluasi, dan akhirnya penyebaran model [3].



Gambar 2.1 :proses CRISP-DM

CRISP-DM memiliki enam proses atau phase yaitu [3]:

1. *Business understanding*

Pada tahap pertama bisa disebut juga tahap pemahaman penelitian, menentukan tujuan proyek penelitian dalam perumusan mendefinisikan masalah data mining.

2. *Data Understanding*

Dilakukan pengumpulan data, kemudian menganalisa data serta evaluasi kualitas data.

3. *Data Preparation*

Persiapkan data mentah kemudian di seting untuk data akhir yang akan digunakan untuk fase selanjutnya, pilih kasus dan variabel yang diinginkan yang digunakan untuk menganalisa sesuai analisa masalah, lakukan transformasi pada variabel tertentu jika diinginkan, bersihkan data untuk alat pemodelan

4. *Modelling*

Pada tahap ini, pilih dan terapkan teknik pemodelan yang tepat, lakukan pengaturan model untuk mengoptimalkan hasil, jika diperlukan lakukan ulang ke tahap persiapan sesuai dengan persyaratan spesifikasi dari teknik data mining tertentu.

5. *Evaluasi*

Melakukan evaluasi satu atau lebih model, tentukan apakah model sudah mencapai tujuan yang diterapkan dalam tahap pertama, mengambil keputusan mengenai penggunaan hasil data mining

6. *Deployment*

Memfaatkan model yang telah dibuat, *deployment* yang sederhana adalah sampai menghasilkan laporan sedangkan *deployment* yang kompleks adalah melaksanakan model untuk proses data mining paralel pada departemen lain.

Fungsi-fungsi dalam data mining mengacu pada Larose (2005) terdapat enam fungsi yaitu [4]:

1. Fungsi deskripsi (*description*)

Fungsi deskripsi adalah cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas. Banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya yaitu Deskripsi Grafis, Deskripsi Lokasi, dan Deskripsi Keragaman.

2. Fungsi estimasi (*estimation*)

Fungsi estimasi adalah fungsi untuk memperkirakan suatu hal yang sudah ada datanya. Fungsi estimasi terdiri dari dua cara yaitu Estimasi Titik dan Estimasi Selang Kepercayaan.

3. Fungsi prediksi (*prediction*)

Fungsi prediksi adalah memperkirakan hasil dari hal yang belum diketahui, untuk mendapatkan hal baru yang akan muncul selanjutnya. Cara memprediksi dalam fungsi ini adalah Regresi Linier.

4. Fungsi klasifikasi (*classification*)

Fungsi klasifikasi atau menggolongkan suatu data. Cara yang digunakan terdiri dari algoritma Mean Vector, algoritma K-nearest Neighbor, algoritma ID3, algoritma C4.5, dan algoritma C5.0

5. Fungsi pengelompokan (*cluster*)

Fungsi pengelompokan, data yang dikelompokkan disebut objek atau catatan yang memiliki kemiripan atribut kemudian dikelompokkan pada kelompok yang berbeda. Algoritma yang digunakan adalah algoritma Hierarchical Clustering, algoritma Partitional Clustering, algoritma Single Linkage, algoritma Complete Linkage, algoritma Average Linkage, algoritma K-Means dan lain-lain.

6. Fungsi asosiasi (*association*)

Fungsi asosiasi adalah untuk menemukan aturan asosiasi (*association rule*) yang mampu mengidentifikasi item-item yang menjadi objek. Algoritma yang digunakan adalah algoritma Generalized Association Rules, Quantitative Association Rule, asynchronous Parallel Mining

Algoritman A Priori termasuk jenis aturan asosiasi pada data mining. Selain a priori, yang termasuk pada golongan ini adalah metode *generalized rule induction* dan algoritman *hash based*. Aturan yang menyatakan asosiasi antara beberapa atribut atau variabel sering disebut *affinity analysis* atau *market basket analysis* [1]. Penting tidak suatu aturan *assosiatif* dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan *assosiatif*.

Aturan *assosiatif* biasanya dinyatakan dalam bentuk :

{siang, media sosial} \rightarrow {facebook} (support = 40%, confidence=50%)

Yang artinya, 50% dari transaksi di database yang memuat item waktu akses “siang” dan jenis web “media sosial” juga memuat item situs “facebook”. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu. Dapat juga diartikan seorang yang mengakses situs di siang hari dan jenisnya media sosial punya kemungkinan 50% untuk mengakses alamat situs facebook.

Analisi asosiasi didefinisi suatu proses untuk menemukan semua aturan *assosiatif* yang mempunyai syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metode dasar analisis asosiasi terbagi menjadi dua tahap :

a. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut [2]:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

Sedangkan nilai *support* dari 2 item diperoleh dari rumus berikut :

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}}$$

b. Pembentukan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif $A \rightarrow B$

Sebagai contoh diambil suatu data transaksi yang di dapat dari penggunaan internet, dengan data transaksi sebgai berikut :

Transaksi	Item yang sesuai
1	Waktu="pagi", jenis ="sosial", alamat web ="www.facebook.com"
2	Waktu="pagi", jenis ="berita", alamat web ="www.detik.com"
3	Waktu="sore", jenis ="niaga", alamat web ="www.tokopedia.com"
4	Waktu="malam", jenis ="niaga", alamat web ="www.olx.com"
5	Waktu="siang", jenis ="berita", alamat web ="www.mivo.tv"

Definisi-definisi yang terdapat pada *assosiation rule*

1. I adalah himpunan yang tengah dibicarakan
Contoh : {pagi, ssosial, www.facebook.com}
2. D adalah himpunan seluruh transaksi yang tengah dibicarakan
Contoh : {Transaksi 1, transaksi 2,, transaksi 5}
3. Proper subset adalah himpunan bagian murni [3]
Contoh :
Ada suatu himpunan $A = \{a,b,c,\}$
Himpunan bagian dari A adalah

Himpunan kosong = $\{\}$

Himpunan 1 unsur = $\{a\}, \{b\}, \{c\}$

Himpunan 2 unsur = $\{a,b\}, \{a,c\}, \{b,c\}$

Himpunan 3 unsur = $\{a,b,c\}$

Proper subsetnya adalah himpunan 1 unsur dan himpunan 2 unsur

4. Item set adalah himpunan item atau item-item di I [3]

Contoh :

Ada suatu himpunan $A = \{a,b,c\}$

Item setnya adalah $\{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}$

5. K-item set adalah item set yang terdiri dari K buah item yang ada pada I . [3]. Intinya K itu adalah jumlah unsur yang terdapat pada suatu himpunan

Contoh :

3 item set adalah yang bersifat 3 unsur

6. Item set frekuensi adalah jumlah transaksi di I yang mengandung jumlah item set tertentu. Intinya jumlah transaksi yang mengandung sejumlah item [5]

7. Frekuensi item set adalah item set yang muncul sekarang-kurangnya “sekian” kali di D . Kata “sekian” biasanya disimbolkan dengan Φ , Φ merupakan batas minimum dalam sebuah transaksi [5]

Contoh :

Pertama kita tentukan $\Phi = 3$, karena jika tidak ditentukan maka frekuensi item set tidak dapat dihitung

8. F_k adalah himpunan semua frekuensi item set yang terdiri dari K item.

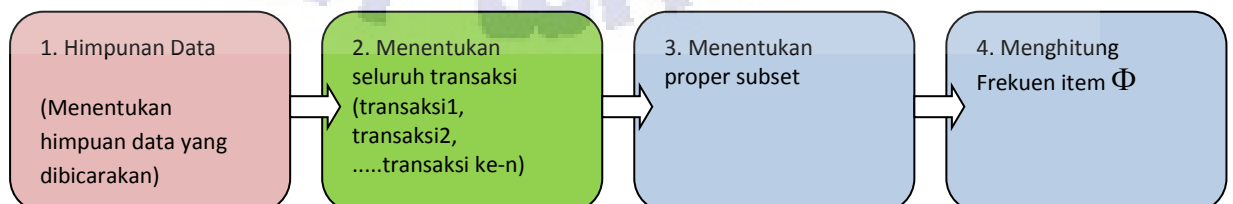
Langkah-langkah algoritma pada *Association Rule* [4]

1. Tentukan Φ
2. Tentukan semua frekuensi item set
3. Untuk setiap frekuensi item set dilakukan hal sebagai berikut :
 - a. ambil sebuah unsur, namakanlah s
 - b. untuk sisanya namakanlah $ss-s$
 - c. masukkanlah unsur-unsur yang telah diumpamakan ke dalam rule if ($ss-s$) then s

Untuk langkah ke-3 lakukan untuk semua unsur.

3. Metode Perancangan

Proses perancangan *A Priori Rule* ini dibutuhkan beberapa tahapan dalam menyusun penelitian. Tahap-tahap yang dilakukan ditunjukkan pada gambar 3.1



Gambar 3.1 : Desain Data Mining

yaitu: (1) Himpunan Data, (2) Menentukan seluruh transaksi, (3) Menentukan *Proper Subset*, (4) Menghitung Frekuensi item

Cara Kerja Algoritma A Priori dalam Penelitian ini

Berikut merupakan cara kerja metode A Priori pada penelitian ini :

1. Persiapan dataset
Dengan menyiapkan data set tabel log dalam bentuk excel, tabel_log.xls
2. Pengelompokan data transaksi pada item yang sesuai
3. Menghitung jumlah item-item yang sesuai pada setiap kolomnya
4. Menentukan rule atau aturan yang berguna untuk menentukan prediksi situs internet yang cenderung muncul

Pengumpulan Data

Teknik pengumpulan data dilakukan dengan mempersiapkan data akses internet bulan Maret 2016. Data yang telah dikumpulkan dari administrasi web server, selanjutnya dilakukan rekap ke microsoft excel sehingga diperoleh data yang valid. Data yang sudah di rekap di microsoft excel siap untuk di olah.

4. Analisa dan Pembahasan

Transaksi akses internet yang ada adalah

Tabel 4.1 : Tabel Akses Internet

No.Trans	tgl.akses	jam. akses	jenis waktu akses	ip address	Alamat Web	kode web	jenis alamat
1	01/03/2016	8:01	Pagi	192.168.1.20	www.facebook.com	1	sosial
2	02/03/2016	8:02	Pagi	192.168.1.21	www.detik.com	2	berita
3	03/03/2016	8:03	pagi	192.168.1.22	www.tokopedia.com	3	niaga
4	04/03/2016	8:04	pagi	192.168.1.23	www.olx.com	4	niaga
5	05/03/2016	8:05	pagi	192.168.1.24	www.mivo.tv	5	berita
6	06/03/2016	8:06	pagi	192.168.1.25	www.toon4.com	6	hiburan
7	07/03/2016	8:07	pagi	192.168.1.26	www.spankwire.com	7	hiburan
8	08/03/2016	8:08	pagi	192.168.1.27	www.fifa.com	8	olahraga
9	09/03/2016	8:09	pagi	192.168.1.28	www.4shared.com	9	pencari
.							
300	11/03/2016	9:26	pagi	192.168.1.25	www.tokopedia.com	3	niaga

Tentukan Item-item pendukung

Tabel 4.2 : Tabel Item Pendukung

Kategori Item	Item pendukung	Simbol	Keterangan
Jenis Waktu	Pagi	A	Akses yang dilakukan periode pagi antara jam 07.00-12.00

	Siang	B	Akses yang dilakukan periode siang antara jam 12.01-15.00
	Sore	C	Akses yang dilakukan periode sore antara jam 15.01-18.00
	Malam	D	Akses yang dilakukan periode malam antara jam 18.01-21.00
Jenis alamat web	Sosial	E	
	Berita	F	
	Niaga	G	
	Hiburan	H	
	Pencarian	I	
	Olahraga	J	
Kode web	Kodeweb 1	K	www.facebook.com
	Kodeweb 2	L	www.detik.com
	Kodeweb 3	M	www.tokopedia.com
	Kodeweb 4	N	www.olx.com
	Kodeweb 5	O	www.mivo.tv
	Kodeweb 6	P	www.toon4.com
	Kodeweb 7	Q	www.spankwire.com
	Kodeweb 8	R	www.fifa.com
	Kodeweb 9	S	www.4shared.com

Buat Tabel kemunculan setiap transaksi

Tabel 4.3 : Tabel Kemunculan Setiap Transaksi

No.T Trans	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
3	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
4	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
5	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
6	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
8	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
9	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
.																			
300	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0

Kemudian hitung jumlah banyaknya kemunculan untuk setiap item

Tabel 4.4 : Tabel Jumlah Kemunculan

No.T Trans	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
3	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
4	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
5	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
6	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
8	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
9	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
.																			
300	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
Σ	79	65	129	27	51	85	89	39	20	16	51	45	45	44	40	24	15	16	20

Kemudian hitung Support dari setiap item :

Tabel 4.5 Nilai Support dari setiap item

NAMA ITEM	JUMLAH	SUPPORT
A	79	26%
B	65	22%
C	129	43%
D	27	9%
E	51	17%
F	85	28%
G	89	29%
H	39	13%
I	20	6%
J	16	5%
K	51	17%
L	45	15%
M	45	15%
N	44	14%
O	40	13%
P	24	8%
Q	15	5%
R	16	5%
S	20	6%

Tentukan Φ

Misalkan kita tentukan $\Phi = 15$, maka kita dapat menentukan frekuensi itemset. Dari tabel diatas diketahui total Φ untuk transaksi $k=1$, semuanya lebih besar dari Φ , maka :

$$F_1 = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}, \{G\}, \{H\}, \{I\}, \{J\}, \{K\}, \{L\}, \{M\}, \{N\}, \{O\}, \{P\}, \{Q\}, \{R\}, \{S\}\}$$

Untuk $k=2$ (2 unsur), diperlukan tabel untuk tiap-tiap pasang item. Himpunan yang mungkin terbentuk adalah :

Tabel 4.6 : Tabel Himpunan yang terbentuk

A,E	A,F	A,G	A,H	A,I	A,J	A,K	A,L	A,M	A,N	A,O	A,P
A,Q	A,R	A,S	B,E	B,F	B,G	B,H	B,I	B,J	B,K	B,L	B,M
B,N	B,O	B,P	B,Q	B,R	B,S	C,E	C,F	C,G	C,H	C,I	C,J
C,K	C,L	C,M	C,N	C,O	C,P	C,Q	C,R	C,S	D,E	D,F	D,G
D,H	D,I	D,J	D,K	D,L	D,M	D,N	D,O	D,P	D,Q	D,R	D,S

Berikut tabel-tabel untuk 2 item set :

T	A	E	f	T	A	F	f	T	A	H	f
1	1	1	P	1	1	1	S	1	1	1	S
2	1	0	S	2	1	0	S	2	1	0	S
3	1	0	S	3	1	0	S	3	1	0	S
4	1	0	S	4	1	0	S	4	1	0	S
5	1	0	S	5	1	0	S	5	1	0	S
6	1	0	S	6	1	0	S	6	1	0	S
7	1	0	S	7	1	0	S	7	1	0	S
8	1	0	S	8	1	0	S	8	1	0	S
9	1	0	S	9	1	0	S	9	1	0	S
.				.				.			
300				300				300			
Σ			13	Σ			20	Σ			12

Dan seterusnya sampai dengan item $\{D,R\}, \{D,S\}$. Angka 13 yang berwarna merah di tabel diperoleh dari data 1 s.ampai dengan 300, dengan menghitung jumlah baris yang item A,E masing-masing bernilai 1 atau yang item F nya bernilai P. Dari keseluruhan tabel yang dibuat yaitu $\{A,E\}, \{A,L\}, \{A,M\}, \{A,N\}, \dots, \{D,S\}$ diperoleh masing-masing jumlah P (ΣP) sebagai berikut :

Tabel 4.7 : Tabel Rekap Jumlah ΣP

ITEM	ΣP	ITEM	ΣP	ITEM	ΣP	ITEM	ΣP
A,E	13	B,E	8	C,E	24	D,E	5
A,F	21	B,F	18	C,F	39	D,F	7
A,G	21	B,G	18	C,G	42	D,G	8
A,H	13	B,H	10	C,H	12	D,H	3
A,I	6	B,I	4	C,I	8	D,I	2
A,J	5	B,J	5	C,J	4	D,J	2
A,K	13	B,K	9	C,K	24	D,K	5
A,L	12	B,L	8	C,L	20	D,L	4
A,M	11	B,M	8	C,M	20	D,M	4
A,N	10	B,N	9	C,N	21	D,N	4
A,O	9	B,O	9	C,O	19	D,O	3
A,P	8	B,P	6	C,P	8	D,P	2
A,Q	6	B,Q	5	C,Q	4	D,Q	1
A,R	6	B,R	5	C,R	4	D,R	2
A,S	6	B,S	4	C,S	8	D,S	2

Cek item yang memenuhi nilai Φ

Dari tabel 4.6, diambil beberapa himpunan yang memenuhi nilai Φ sebesar 15 diantaranya :

$$F2 = \{ \{A,F\}, \{A,G\}, \{B,F\}, \{B,G\}, \{C,E\}, \{C,F\}, \{C,G\}, \{C,K\}, \{C,L\}, \{C,M\}, \{C,N\}, \{C,O\} \}$$

Kombinasi dari itemset dalam $F2$, dapat kita gabungkan menjadi calon 3-itemset. Itemset-itemset yang dapat digabungkan adalah itemset-itemset yang memiliki kesamaan dalam k-1 item pertama.

Untuk $k=3$ (3 unsur), diperlukan tabel untuk tiap-tiap pasang item. Himpunan yang mungkin terbentuk adalah :

$$\{A,F,G\}, \{B,F,G\}, \{C,E,F\}, \{C,E,G\}, \{C,E,K\}, \{C,E,L\}, \{C,E,M\}, \{C,E,N\}, \{C,E,O\}, \{C,F,G\}, \{C,F,K\}, \{C,F,L\}, \{C,F,M\}, \{C,F,N\}, \{C,F,O\}, \{C,G,K\}, \{C,G,L\}, \{C,G,M\}, \{C,G,N\}, \{C,G,O\}, \{C,K,L\}, \{C,K,M\}, \{C,K,N\}, \{C,K,O\}, \{C,L,M\}, \{C,L,N\}, \{C,L,O\}, \{C,M,N\}, \{C,M,O\}, \{C,N,O\}$$

Setelah mendapat himpunan 3 unsur yang terbentuk maka hitung kemunculannya, dari hasil tersebut kita akan ambil yang memenuhi Φ yang sudah kita tentukan.

Tabel 4.8: table-tabel 3 item set

T	A	f	G	F	T	B	F	G	F	T	C	E	K	F
1	1	1	0	S	1	1	1	0	S	1	1	0	2	1
2	1	0	0	S	2	1	0	0	S	1	0	0	1	1
3	1	0	0	S	3	1	0	0	S	1	0	0	1	1
4	1	0	1	S	4	1	0	0	S	1	0	0	1	1

5	1	0	0	S		5	1	0	1	S		1	0	0	1	1
6	1	0	0	S		6	1	0	0	S		1	0	1	2	1
7	1	0	0	S		7	1	0	0	S		1	0	0	1	1
8	1	0	0	S		8	1	0	0	S		1	0	0	1	1
9	1	0	0	S		9	1	0	0	S		1	0	0	1	1
.																
300																
					Σ	0						Σ	0			
														Σ	24	

Dan seterusnya sampai dengan item {C,MO},{C,N,O}. Angka 24 yang berwarna merah di tabel diperoleh dari data 1 s.ampai dengan 300, dengan menghitung jumlah baris yang item C,E,K masing-masing bernilai 1 atau yang item F nya bernilai P

Tabel 4.9 : Tabel Rekap Jumlah Σ P

ITEM	Σ P	ITEM	Σ P	ITEM	Σ P	ITEM	Σ P	ITEM	Σ P	ITEM	Σ P
A,F,G	0	C,E,L	0	C,F,K	0	C,G,K	0	C,K,L	0	C,L,N	0
B,F,G	0	C,E,M	0	C,F,L	20	C,G,L	0	C,K,M	0	C,L,O	0
C,E,F	0	C,E,N	0	C,F,M	0	C,G,M	21	C,K,N	0	C,M,N	0
C,E,G	0	C,E,O	0	C,F,N	0	C,G,N	21	C,K,O	0	C,M,O	0
C,E,K	24	C,E,G	0	C,F,O	19	C,G,O	0	C,L,M	0	C,N,O	0

Dari table 4.9, d idapat F3 yang memenuhi $\Phi = 15$ yaitu :

$$F3 = \{\{C,E,K\}, \{C,F,L\}, \{C,F,O\}, \{C,G,M\}, \{C,G,N\}\}$$

Buat *Rule* atau Pola

Berdasarkan himpunan yang memenuhi nilai $\Phi = 15$, maka dapat dibuat rule sebagai berikut

Tabel 4.10 : Tabel Rule

No.	Himpunan	Antecedent	Consequent	Rule	Keterangan
1	{C,E,K}	C,E	K	If C and E Then K	Jika Waktu = Sore dan Jenisalamamat=Sosial maka cenderung mengakses www.facebook.com
2	{C,F,L}	C,F	L	If C and F Then L	Jika Waktu = Sore dan Jenisalamamat=Berita maka cenderung mengakses www.detik.com
3	{C,F,O}	C,F	O	If C and F Then O	Jika Waktu = Sore dan Jenisalamamat=Berita maka cenderung mengakses www.mivo.com
4	{C,G,M}	C,G	M	If C and G Then M	Jika Waktu = Sore dan Jenisalamamat=Niaga maka cenderung mengakses www.tokopedia.com

5	{C,G,N}	C,G	N	If C and G Then N	Jika Waktu = Sore dan Jenisalamamat=Niaga maka cenderung mengakses www.olx.com
---	---------	-----	---	----------------------	--

Hitung *Support* dan *Confidence*

$$\text{Support} = \frac{\Sigma \text{ jumlah rule yang mengandung anteseden}}{\Sigma \text{ jumlah rule}} \times 100\%$$

$$\text{Confidence} = \frac{\Sigma \text{ jumlah rule yang mengandung qonsekuen}}{\Sigma \text{ jumlah qonsekuen pada semua rule}} \times 100\%$$

Tabel 4.11 : Tabel *Support* dan *Confidence*

If antecedent then consequent	Support	Confidence
If C and E Then K	$(1/5) \times 100\% = 20\%$	$(1/5) \times 100\% = 20\%$
If C and F Then L	$(2/5) \times 100\% = 40\%$	$(1/5) \times 100\% = 20\%$
If C and F Then O	$(2/5) \times 100\% = 40\%$	$(1/5) \times 100\% = 20\%$
If C and G Then M	$(2/5) \times 100\% = 40\%$	$(1/5) \times 100\% = 20\%$
If C and G Then N	$(2/5) \times 100\% = 40\%$	$(1/5) \times 100\% = 20\%$

Lakukan perkalian antara *support* dan *confidence*

Tabel 4.12 : Tabel Perkalian *Support* dan *Confidence*

No.	If antecedent then consequent	Support	Confidence	Support x Confidence
1	If C and E Then K	20%	20%	0.004
13	If C and F Then L	40%	20%	0.008
14	If C and F Then O	40%	20%	0.008
15	If C and G Then	14%	40%	0.008

	<i>M</i>			
16	<i>If C and G Then N</i>	7%	40%	0.008

Setelah didapat hasil perkalian antara support dan confidence, pilihlah yang hasil perkaliannya paling besar. Hasil paling besar dari perkalian perkalian tersebut merupakan rule yang dipakai pada saat untuk meunjukkan perilaku pengguna internet

Tabel 4.13 : Tabel Perkalian Support dan Confidence

No.	<i>If antecedent then consequent</i>	<i>Support</i>	<i>Confidence</i>	<i>Support x Confidence</i>
8	<i>If C and F Then L</i>	40%	20%	0.008
9	<i>If C and F Then O</i>	40%	20%	0.008
10	<i>If C and G Then M</i>	40%	20%	0.008
11	<i>If C and G Then N</i>	40%	20%	0.008

5. Kesimpulan

Proses penentuan pola melihat aktifitas pengguna internet dapat dilakukan dengan menerapkan data mining dengan metode algoritma *apriori*. dengan metode tersebut penentuan pola dapat dilakukan dengan melihat hasil dari kecenderungan tiap orang mengakses *web* berdasarkan kombinasi 3 itemset. Berdasarkan tabel 4.13 didapatkan rule sebanyak 4 yang dapat dijadikan untuk melihat aktivitas penggunaan internet. Pemakaian ke -4 rule tersebut akan mempermudah pengelolaan internet yaitu pengaturan *bandwitdh* pada setiap alamat web yang kurang dominan dan penambahan *bandwidth* pada alamat web yang dominan dengan unsur penunjang jenis waktu akses (pagi, siang, sore atau malam) serta jenis alamat web yang diakses (berita, hiburan, pencari, sosial, niaga). Pembagian *bandwidth* dilakukan agar internet dapat berjalan optimal bagi semua pengakses.

6. DAFTAR PUSTAKA

- [1] Kennedy Tampubolon (2013), **Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan**
- [2] Robi Yanto (2015), **Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat**
- [3] Heruandika Cahyono Pratama (2014), **Penerapan Algoritma Apriori Dalam Menemukan Hubungan Data Awal Masuk Mahasiswa Dengan Prestasi Akademik (Studi Kasus : Stai Miftahul Ulum Tanjungpinang)**
- [4] Witri Widiawati (2015), **Penerapan Algoritma Apriori Dengan Analisa Keranjang Belanja Untuk Sistem Persediaan Alat-Alat Kesehatan**
- [5] Pria Nita Utari (2015), **Penerapan Metode Association Rule Menggunakan Algoritma Apriori Untuk Analisa Pola Data Kecelakaan Pesawat Dari Tahun 1967-2014 Di Indonesia**
- [6] Roger Pressman, **“Rekayasa Perangkat Lunak”**, Andi Offset, 2009